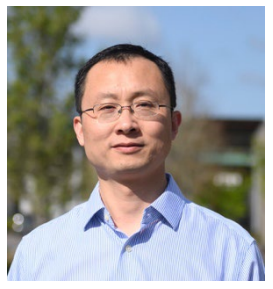
中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science, CAS

生物信息学研究中心

Center of Bioinformatics

学术报告



题　目： Supervised learning of phenotype-associated subpopulations from millions of cells

报告人： **Prof. Zheng Xia**

Biomedical Engineering Department, Knight Cancer Institute, Oregon Health & Science University

时　间： 7 月 28 日（星期五）上午 09:30

地　点： 数学院南楼 620 室（#腾讯会议：296-866-505）

摘　要： Single-cell sequencing (scSeq) technologies are revolutionizing biomedical research and clinical practice by enabling the comprehensive characterization of cells from complex tissues. scSeq can identify cell types, states, and lineages of different cell subpopulations associated with diseases in a heterogeneous tissue ecosystem, providing a breakthrough opportunity to understand biological and clinical mechanisms. However, interpreting complex single-cell data from highly heterogeneous cell populations remains challenging. Identifying cell subpopulations that drive phenotypes, such as tumor metastasis, treatment resistance, and survival outcome, is of indispensable importance since it will facilitate cell-type targeted therapies and prognostic biomarker discovery. Clustering-based methods depend on the subjective clustering step and are often suboptimal since unsupervised clustering may not capture phenotype-specific subpopulations.

To overcome this challenge, we recently developed a novel framework called SCISSOR, which identifies phenotype-associated subpopulations from single-cell assays by leveraging the wealth of phenotype and bulk-omics datasets (Sun et al., Nature Biotechnology, 2022). Thus, this tool paves the way for employing widely available bulk phenotype information to unlock the disease-relevant subpopulations for cellular target discoveries from scSeq data. Moreover, an increasing number of single-cell experiments are designed to profile multiple samples from different conditions, such as treatment resistance versus responder groups. Identifying subpopulations unique to each phenotype will improve phenotype-specific gene signal detection to facilitate reliable downstream analysis. Since clustering-based methods are often suboptimal, we propose to use a novel ‘learning with rejection’ platform to learn high confidence phenotype-enriched subpopulations from millions of cells (Ren et al., Nature Machine Intelligence, 2023).

报告人简介： Dr. Zheng Xia is an Associate Professor in the Department of Biomedical Engineering, as well as a member of the Computational Biology program and Knight Cancer Institute at Oregon Health & Science University. He obtained a BS in Automation in 2003 and a PhD in Pattern Recognition and Intelligent Systems in 2009, both from Zhejiang University. His lab develops bioinformatics tools for next-generation sequencing data analysis and machine learning algorithms for large-scale biomedical data interpretation. In collaboration with biologists and clinicians, he has used big data bioinformatics analysis to gain novel biological and clinical insights into aging, stem cells, neurological disorders, and cancers. More than 70 peer-reviewed papers have been published through his methodology development and extensive collaborations, including 20 in Nature, Cell, and Science series. His current research interests include developing single-cell data analysis algorithms (Nature Biotechnology 2018, 2022; Nature Machine Intelligence 2023), investigating cancer metastasis and treatment resistance (Nature Communications 2021, 2022; PNAS 2020; Clinical Cancer Research 2020, 2021), and advancing cancer immunotherapy (Nature 2022).